DOCUMENT RESUME

ED 208 088 UD 021 693

AUTHOR Deck, Dennis; Arter, Judy
TITLE Evaluation Designs for Title I Summer Programs.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
PUB DATE Apr 81
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Measurement; *Program Evaluation; Scores; *Summer Programs; Test Construction
IDENTIFIERS Elementary Secondary Education Act Title I

ABSTRACT

This paper presents key concepts and practical suggestions to use in evaluating three types of Title I summer programs. A description of the summer program characteristics is followed by outlines of the three evaluation designs: (1) norm referenced design; (2) criterion referenced design; and (3) comparison group design. The techniques, procedural guidelines, requirements for use, and advantages and disadvantages of each design are reviewed. The paper asserts that these evaluation designs are easy to implement and would yield reasonably valid conclusions. A discussion of quality control, accuracy in score measurement, and the relationship between test content and instruction concludes the paper. (APM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208088

Evaluation Designs for Title I Summer Programs

by
Dennis Deck
Judy Arter

Northwest Regional Educational Laboratory
Title I Technical Assistance Center
710 S.W. Second Avenue
Portland, OR 97204

Paper presented at the annual meeting of the
American Educational Research Association,
Los Angeles, California
April 13-17, 1981

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. Margaret Rogers
N.W. Regional Educational Lab.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

UD021693

Much has been written about evaluating the academic outcomes of Title I and other compensatory programs, but little attention has been given to the unique problems of evaluating summer programs. Evaluation designs developed for use with regular term Title I programs may be inappropriate or must be adapted for use with summer programs. Perhaps because the Title I regulations are more flexible and there is no school schedule to work around, there is great variation in these programs. The short term nature and limited focus of summer programs creates special difficulties for the evaluator. The characteristics of a particular program can make certain evaluation designs are more appropriate than others.

Since randomized studies are rarely possible, the evaluator must be able to select a design that controls for alternative explanations of the results. The purpose of this paper is to suggest evaluation designs that seem appropriate for summer Title I programs and procedures for implementing these designs. We will describe various dimensions of summer programs which affect the choice of an evaluation design, and we will discuss threats to the validity of summer program evaluations that are particularly troublesome.

Since this paper presents an overview, it does not provide the procedural or technical detail necessary to full implement the recommended designs, but references are provided for further information. We recognize that most districts do not have staff trained in evaluation methodology and have tried to focus on key concepts and practical suggestions.

Characteristics of Summer Programs

After reviewing program descriptions and evaluations of summer Title I programs from several western states, we have identified sources of variation in summer programs that are critical in planning an evaluation of academic achievement. Before a school district adopts a particular evaluation design, careful consideration of how the nature of the program affects the outcomes of the evaluation is necessary.

Instruction. A summer program may last anywhere from two to eight weeks. During that time instruction may be given as much as six hours a day or as little as a couple of hours a week. This means that while some programs cover a broad range of objectives, others focus on a few specific objectives. This range has several implications for selecting a measure of impact and an evaluation design.

First, it seems unreasonable to expect to detect growth on a standardized achievement test unless the program has a broad focus and rather intensive instruction. Most school districts currently use standardized, norm-referenced achievement tests which sample only a small number of items in any skill area. With a program having a narrow focus, it is possible that only one or two items cover what was taught. The importance of selecting a test that will be sensitive to the instruction provided by the program cannot be stressed too much.

Second, the limited time available for instruction in some programs preclude spending much time in testing or conducting the evaluation. In

some situations, spring and fall test data can be used rather than administering additional tests. In other situations, tests that are used for student monitoring can be used.

Individualization. While many Title I programs attempt to individualize instruction, the degree of individualization differs considerably. Sometimes this means individual help while working on exercises. In other programs each student sets his own pace working through the same materials. In still others a unique set of activities is prescribed for each student. Often, individualized work is interspersed with group activities.

One implication for program evaluation relates to the selection of a test to measure growth. The extent to which each student receives a very different treatment determines whether a single pool of items could be considered an adequate measure of achievement for the group. While a broad ranged test may be appropriate, often a criterion-referenced measure tailored to each student would be preferable.

Nonacademic objectives. Many summer programs can be distinguished by their emphasis on affective outcomes. While there is instruction in basic skills, high interest materials and supportive activities are used to improve the students' self esteem or attitudes toward school. In such a program it is appropriate to supplement or replace achievement measures with attitudinal scales, observation, or other measures. One would expect such a program to improve student achievement but perhaps not immediately.

Student selection criteria. The procedure for selecting students into a program is a very important consideration in planning an evaluation. Like other Title I programs, summer projects generally use an informal process based on teacher referral and perhaps a general cutoff score on a standardized test. The referral, however, might be based on such criterion as the student needing assistance in a particular skill area, in building self-esteem, or in general skill development. Generally, the students selected can elect not to attend. Often the participants are primarily students who participated in the regular term program but sometimes the criteria for determining a "needy" student are greatly relaxed so that the district can be sure to fill all slots in the program.

One implication of the selection method is that if the pretest scores are used to select students, regression to the mean could confound the treatment effect in some designs by making the program look more effective than it actually was. Another implication is that, unlike regular term programs, there is often an excellent opportunity to select a comparison group of students who are eligible for Title I but who do not participate. Such a comparison group would have to be selected using the same criteria as the participating students.

Other instruction. Since the evaluator wants to distinguish between growth due to the summer program and growth due to other factors, it is important to consider influences outside the program. No other formal instruction is likely during the summer so it may seem reasonable to

expect the same or worse performance at posttest time without the summer program. However, the student still reads and applies math skills throughout the summer. Growth should be expected on certain skills due to practice, informal instruction by parents, or maturation. It appears that "summer dropoff" involves slowing but not stopping the student's rate of growth, at least as measured with standardized achievement tests administered spring and fall (Roberts, 1980; Stenner & Bland, 1979). At any rate, a comparison group or other controls may be necessary to distinguish growth attributable to the program.

Relationship to regular term program. The similarities in the objectives, materials, and students between certain summer programs and their regular term counterparts opens up the possibility of integrating the evaluations of the two components. When the summer activities are simply an extension of the regular term, there may be little value in evaluating them separately.

Evaluation Designs for Summer Programs

Educators have a long tradition of using both a pretest and posttest to determine how much growth occured during the educational program. However, since many of the evaluations conducted have failed to estimate how much would have been learned without the program, the results of these studies have often been uninterpretable.

Designing an evaluation that will yield interpretable results is like the criminal lawyer preparing a court case. He develops a theory of how the crime took place by looking for evidence that eliminates as many alternative explanations as possible. In program evaluation, the evaluator makes assumptions that lead to estimates of how students would have performed without the program, and collects data in a way that makes alternative explanations of the results less plausible. The effects of the program must be separated from that of maturation, other classroom instruction, informal instruction in the home, testing problems, characteristics of the particular group selected, and so on.

Three evaluation designs are suggested here. Each makes different assumptions about student growth that allow the evaluator to estimate how students would have performed without the Title I instruction. Because each assumption may be tenable only for summer programs having certain characteristics, none of the designs are universally applicable.

In describing each design, we have tried to suggest procedural guidelines and potential variations. We suggest what types of programs can best be evaluated by each design and listed advantages and disadvantages of each design.

Norm-Referenced Design

The norm-referenced design compares the growth of the Title I students with the growth that would have been expected without the Title I instruction. This no-treatment expectation is determined by the posttest performance of students in the norming sample of a standardized test

whose pretest status was the same as the Title I group. At posttest time the expected achievement is subtracted from the actual achievement to determine the effect of the program. The design is used essentially the same way as in evaluating regular term Title I programs (Tallmadge & Wood, 1980).

Requirements for Use

The norm-referenced design seems best suited to summer programs which provide intensive instruction on a broad range of objectives. Tests that are normed typically sample a very small number of items from each skill area so that achievement of a wide range of objectives can be assessed. It is unlikely that such a test would be sensitive to instruction that focused on only a few objectives or that was quite short in duration. The Title I program should provide three or more hours a day for six to eight weeks.

The basic assumption of the norm-referenced model is that without treatment students will tend to remain at the same percentile rank. (Note, however, that since the scores of an individual student fluctuate too much, the assumption is based on the average of the group.) If students in the local population are not similar to those in the norming population or if the school curriculum differs from those schools included in the sample, this "equipercentile" assumption may not hold and the estimate of treatment effect may not be accurate.

There are several ways to check this assumption. Check the description of the norming sample in the technical manuals of the test to make sure that students like those in the local population were included in the norming. If historical data are available on the local students, one could check to see if, on the average, students in the district maintain their percentile ranks prior to Title I instruction.

For evaluating summer programs, the design also assumes that little or no supplementary instruction occurred between pretest and posttest other than the summer program. Since the test must generally be given in the spring and fall at norm dates, some students might receive several weeks of regular term Title I instruction between the pretest and posttest, thus confounding the summer evaluation. On the other hand, if the regular term program starts late and ends early in the school year, there may not be a problem.

The design is also well suited for the program which extends the materials and objectives of the regular term Title I program with the same students. In some cases it may not be worthwhile to evaluate the two programs separately so that the focus of the evaluation would be their combined effects. If many regular term students do not participate during the summer, gains could be computed separately for the regular term only and the regular plus summer participants.

Procedural Guidelines

The guidelines for the norm-referenced design are detailed in the User's Guide (Tallmadge & Wood, 1980) for the Title I Evaluation and Reporting System. The following are some highlights as they apply to summer programs.

Select an appropriate standardized test. The test must have empirical national or local norms for the spring and fall. These norms should be based on a large sample of students which is representative of the local group of students. The test should reflect what is taught in the summer program and should contain few objectives that are not covered. It would be difficult to find a suitable norm-referenced test for short summer programs that focus on a limited set of objectives.

Select students. Students most in need of supplementary instruction should be selected for the program. This may be done using a variety of methods as long as the pretest scores were not used in any way in the selection process. This is to avoid regression to the mean. Since participation in the summer program is usually voluntary, statistical adjustments for regression that require a strict pretest cutoff for selection cannot be used.

Pretest and posttest students. The spring pretest and fall posttest should be administered within two weeks of the empirical norm dates of the test and the instructions for administering the test must be carefully followed. If the test cannot be administered close to the norm dates but within six weeks, it is possible to interpolate the norms. Students should be tested at their functional level since the test level recommended for each grade level may be too difficult for Title I students. An attempt should be made to use the form of the test used by the publisher in the norming study. When possible, the same form and level of the test should be used for both pretest and posttest.

Compute the treatment effect. Only the data from students having both a pretest and posttest score are used in the analysis. The pretest and posttest scores are converted to NCEs or expanded standard scores and averaged. The effect of the program is the average posttest NCE minus the average pretest NCE.

Advantages and Disadvantages

The primary advantage of the norm-referenced design is its simplicity and familiarity to school district staff. The procedures are not difficult to implement and should be familiar to those who have evaluated regular term programs. Another advantage is that no additional testing may be required if the district has already administered an appropriate test as part of a district testing program or other evaluation requirements.

The design has several disadvantages. First, the assumption that the Title I students would achieve at the same rate as students in the norming sample having the same initial status may not be accurate for the local population. Second, regular term Title I instruction that occurs after the spring pretest or before the fall posttest will bias the estimate of the gains. Third, the rather long time span between pretest

and posttest allows many opportunities for learning (and forgetting) to occur due to factors other than the program. Fourth, the test selected will probably be sensitive to only rather intensive summer instruction.

Criterion-Referenced Design

A criterion-referenced design determines the progress of Title I students with respect to a specified criterion or standard. The criteria or standard may be defined in terms of some minimum number of test items answered correctly or in terms of some level of performance. In either case the measure of achievement is referenced directly to well defined instructional objectives. The effect of the program is determined by comparing the posttest performance of students with their pretest performance.

Requirements for Use

The criterion-referenced design seems best suited to summer programs which are based on carefully identified objectives, especially when very different objectives are prescribed for each student. Often these will be very short programs which provide help to students in a fairly limited number of skill areas or objectives.

This design is essentially a pre-post design on a single group of students and lacks a procedure for determining whether the observed growth is due to the program or to other factors. The assumption is that in the absence of the Title I instruction, students would not have improved on the posttest. This seems like a rather strong assumption, yet there do seem to be situations in which the assumption might be reasonable. Unlike regular term programs, the student is not receiving any formal instruction in addition to Title I. If there is only a short time interval between the pretest and posttest, it is unlikely that much growth would be expected due to maturation, informal instruction in the home, television programs, or similar influences. Also, the development of some skills seems to be more dependent on direct instruction than others. For example, improvement on a specific list of vocabulary words can be more easily attributed to the program than can improvements in general comprehension objectives. Mathematic tests appear to be more sensitive than reading tests. A large district might be able to plan a study to determine whether any growth tends to occur during the interval between pretest and posttest without direct instruction.

One point of clarification would be made at this point. The criterion-referenced <u>design</u> is not equivalent to criterion-referenced <u>testing</u>. The design may incorporate criterion-referenced tests but those same tests could be used in either of the other designs as well.

Procedural Guidelines

The criterion-referenced model is very flexible and there are many possible variations in the way the model might be applied. Some basic guidelines are:

Identify clear instructional objectives. The critical feature of this design is that student achievement is referenced to a domain or set of behaviors. If the objectives are clearly defined, one can easily distinguish between test items or tasks that do or do not measure that objective. Knowing the student's performance on the test, one will know which domain or skill areas the student has mastered and which areas are weak. Some types of objectives (see Nitko, 1980, for a review) are:

1. Specific skill or knowledge such as "addition of three place numbers with carrying"

2. A continuum of skill complexity like "addition of two single digit numbers without carrying" through to "addition of three 3-digit numbers with carrying"

3. Proficiency from novice to expert in a skill such as composition

Objectives can be developed either before or after students are selected into the program. If after, then objectives and criteria for successful completion of the objectives could be individualized.

Select a sample of items to measure each objective or skill domain. A valid and reliable measure of each objective or skill domain to be evaluated is needed. The options include:

1. A commercially available criterion-referenced or diagnostic test that matches the curriculum

2. A test developed to parallel the program materials

3. A customized test developed from an item bank or by a district testing office

4. A skills checklist that can be administered reliably

Some guidelines for developing or evaluating criterion-referenced tests are available (e.g., Hambleton & Eignor, 1978; Popham, 1978), but the field is complicated by the diversity in types of these tests.

Establish Performance Criteria

If a mastery approach is taken, a performance criterion should be set up for each skill so that one can determine when an objective has been reached. These criteria could be in relation to one of the following:

1. Number of items correct on each cluster of items

2. Proportion of individually prescribed objectives mastered

3. Proportion of students mastering each objective or meeting each criterion

4. Score indicating degree of proficiency, or level of task complexity

Select program participants. The most needy students are to be selected into the summer program. This can be done using a variety of procedures as long as the pretest scores are not used in any way in the selection process. Other test scores, teacher observation, progress through a reading series, and other measures can be used.

Pretest and posttest participating students.The nature of the objectives and criteria for success determine when pre and posttesting can occur. In some cases, the same measure is given just before and just after the relevant unit of instruction. In other cases testing would occur at the beginning and end of the program. Note that the test items completed may be different for each student if each student receives instruction on different objectives.

Compute the growth in achievement. The difference in achievement between the pretest and posttest can be expressed in the manner specified by the criteria for success established above.

Advantages and Disadvantages

The main advantage of the design is its flexibility. The design can be adapted to many types of programs. Suitable measures for student monitoring may already be incorporated in the program. There are few constraints on when testing should take place and there is no comparison group. Another advantage is that the design can yield timely information that is tied directly to instruction.

A major disadvantage of the design is that the design is weak. Differences in performance may be due to causes other than the Title I program. Its flexibility may conceal measurement problems or ill-defined objectives. If the design is not carefully integrated into the instructional program, much time can be wasted in testing and recordkeeping. The test information will be most useful if tied directly to the curriculum.

Comparison Group Design

The comparison group design is implemented by establishing two groups of students that are similar in all respects except that one group participates in the summer program and the other does not. Both groups are pretested and posttested under the same conditions and at the same time. The relative progress of the treatment group over the comparison group yields an estimate of the effect of the program. The design is used in the same way as for evaluating other Title I programs (Tallmadge & Wood, 1980).

Requirements for Use

The assumption behind the design is that the comparison group is an adequate control for alternative explanations of the evaluation results so that any differences in the progress of the groups over the summer can be attributed to the Title I program. In order to ensure the adequacy of this control the two groups need to be very similar on all educationally important variables such as, race, gender, socioeconomic status, and pretest status.

The ideal way to ensure equivalent groups is to randomly assign students to the treatment and no-treatment conditions. Since randomization is rarely possible in field settings approximations to true randomized selection must be used. The best way to proceed is to select a comparison group that can be logically assumed to be equivalent to the Title I group and then compare the two groups on educationally relevant variables to establish that, in fact, they are similar. Finding a comparison group for a summer program probably is easier than finding one for a regular term program.

Procedure Guidelines

The User's Guide (Tallmadge & Fagan, 1980) describes the procedure in more detail, but some highlights that pertain to summer programs follow.

Select a test. The test chosen should measure the summer program objectives. The test used need not have norms. However, the test chosen must still be reliable and valid for measuring progress of students.

Select Title I and comparison students. Students in the treatment and comparison groups must be selected in the same manner in order to avoid differential regression between the groups. Some options for selecting a comparison group are:

- Select a pool of students eligible for the summer program. Those students who do not elect to participate would serve as the comparison group. Consider whether the elective nature of this selection method produces a nonparticipating group that is different in educationally relevant ways from the Title I group.

- The comparison group students might be selected from another school not participating in the summer program using the same objective procedure objective criterion used in the participating school.

Students must be selected who have the same educational experiences between the pretest and posttest as the treatment group except participation in the summer program. Since testing must generally occur in the spring and fall, consideration must be given to whether students are receiving any different regular term Title I instruction during that period. Check the similarity of the Title I and comparison groups.

Pre and posttest. The treatment and comparison students must be tested at the same time under the same conditions. Testing need not occur on the empirical norm dates if a norm-referenced test is used. Since it is rarely possible to test the comparison group students during the summer, testing must usually occur during the spring and fall.

Compute the results. Include in the analyses only those treatment and comparison students that have both pre and posttest scores. The decision of what analysis to use is complex and controversial. Professional judgment is needed. When students are randomly assigned (or when assignment is random in effect) it is possible to use analysis of

covariance to adjust for any random differences between the pretest status of the two groups. When groups are assigned based on stable group difference such as volunteering for the program, as is typical in this application, Tallmadge & Wood (1980) suggest a principle axis adjustment (see also Kenny, 1975).

Advantages and Disadvantages

The major advantage to this design is that it provides the best estimate of the effect of the summer program on achievement, if an adequate comparison group is used. Finding an adequate comparison group is easier for summer programs than for other Title I programs. Regular testing data can also usually be used, so that no extra testing is required.

The major disadvantage is that the analysis of the data requires more sophistication if statistical adjustments are used. Also, it may be difficult to obtain a comparison group unless testing is done in the spring and fall.

Discussion

In this paper we have proposed three evaluation designs that seem appropriate for summer Title I programs. We included only designs which seemed relatively easy to implement and which would yield reasonably valid conclusions when implemented properly. There will be situations, however, when none of the suggested designs can be implemented without violating assumptions or guidelines. For example, a brief program with poorly defined objectives and no suitable testing materials would have difficulty implementing the criterion-referenced or norm-referenced design. The summer program using a norm-referenced test for which the empirical norm dates overlap with the regular term program will confound the effects of the two programs.

Regardless of the design selected to evaluate summer programs, there are several issues that should be considered before implementing the evaluation.

Match between content of test and instruction. The match between what is taught and what is assessed is a very important feature of any evaluation. While this may seem intuitively obvious, it is probably the evaluation guideline most frequently disregarded. Too often, a summer program with a definite focus will be evaluated using a broad standardized achievement test. A total reading score, for instance, should not be considered a reasonable measure of a student achievement in a 2-week summer program working exclusively on vowel sounds. It is unlikely that more than a couple of the test items would reflect what was taught.

If a test is an appropriate tool to evaluate a particular program, at least four conditions should hold: (1) the test should measure most of the instructional objectives; (2) the number of items in each skill area should be roughly proportioned to the relative emphasis of that skill in instruction; (3) there should be few items on the test that measure objectives that were not covered during the program, and (4) the tasks or

item formats should be familiar to the students. Each of these criteria should be applied in evaluating the test or subtest selected for the evaluation.

Number of students. To base decisions on evaluation results, one would like to be confident that the observed evaluation results are due to actual program effects rather than random fluctuation due to measurement error or to the particular sample of students participating. Unless data from a substantial number of students were used in computing achievement gains, error can easily obscure program effects. Since few school districts have the resources to apply statistical significance tests to their evaluation data, the evaluation results are often used without regard to the number of students participating in the program.

Typically, summer programs serve only a small number of students. Even if there is little turnover for the duration of the project, estimates of program impact will generally be based on a limited number of pre- and posttest scores.

Since the number of students served by a program cannot be increased simply to improve the evaluation, other methods must be sought to increase the stability of the results. The accepted methods are to follow through in make-up testing, to aggregate results, and to watch for trends.

When the scores of only a few program participants are available, substantial improvements in the stability of group means can be realized by ensuring that both pre- and posttest scores will be available from as many students as possible. This requires make-up testing It may also require coordination with other schools to obtain scores of students who have transferred within the district or graduated.

Aggregating scores across school buildings, across years of the program, or across grades is a very effective method of increasing the stability of evaluation results. When the program has been implemented in a similar way across one or more of these dimensions, better estimates of program effectiveness can be computed from the combined scores. For example, a district offering a small program of five students in both third and fourth grades, might average the scores from both grades and from two years of the program, yielding a single gain based on about 20 students. Care must be taken, though, to combine scores from different grades only if NCEs are used. Care must also be taken in interpreting the results when such combinations are used.

Quality control. The effects of summer Title I programs that can be detected with any of the evaluation designs is likely to be small. If the evaluation is not done carefully, the resulting error can obscure the actual treatment effect. Experience with evaluations of regular term Title I programs suggests that some of the more common sources of error are:

- Failure to follow the guidelines for the evaluation design
- Lack of match between the content of the test and what was taught

- Improper test administration
- Clerical errors in recordkeeping, scoring, and score conversions

Burden of the evaluation. Due to the brief nature of many summer programs, teachers are understandably reluctant to use instructional time for testing students or planning time for scoring those tests. We have suggested above several ways to make maximum use of scores that may have been collected for other purposes. It may be possible to use spring and fall scores from a district-wide testing program or from the evaluation of the regular term Title I program for the evaluation. The norm-referenced design would be implemented most efficiently in this way. Student monitoring during the program using tests referenced to the curriculum can often be used for evaluation purposes, particularly with the criterion-referenced design.

Other forms of evaluation. Measuring the academic achievement of summer Title I students is not the only form of evaluation that can provide useful information about the effectiveness of the program. It may not even be the most efficient or useful for this type of program. There are other outcomes to consider, particularly with programs which emphasize attitudes or self concept. Process or implementation evaluations can provide information about how well the program is functioning from an other perspective or about the extent to which the program was implemented. Often, the results of such studies translate more easily into the program decisions than application of the designs discussed here.

References

Hambleton, R.K., & Eignor, D.R. Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 1978, 15, 321-327.

Kenny, D.A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin, 1975, 82, 345-362.

Nitko, A.J. Distinguishing the many varieties of criterion-referenced tests. Review of Educational Research, 1980, 50, 461-485.

Popham, W.J. Criterion-referenced measurement, Englewood Cliffs, NJ: Prentice-Hall, 1978.

Roberts, S.J. Cognitive growth over the summer. RMC Research Corporation, February, 1980.

Stenner, A.J., & Bland, J.D. Assumptions about summer growth implicit in norm-referenced achievement tests. Paper presented to the annual meeting of the American Educational Research Association, San Francisco, CA, April, 1979.

Tallmadge, G.K., & Wood, C.T. User's Guide--ESEA Title I evaluation and reporting system. RMC Research Corporation, Mountain View, CA, 1980.